

AAAAATAAAATCAT ATG AAA AAT ATT AAA K
 AAA TTA TTA AAA GAA TAT AAA TCA CAA TTA ATT GAA TTA ATT GAA CAA TTT GAA GCA
 GGT ATT GGT TTA ATT TTA GGA GAT GCT TAT ATT CGT AGT CGT GAT GAA GGT AAA ACT TAT
 TGT ATG CAA TTT CAC TCC AAA AAT AAG GCA TAC ATG GAT CAT GIA TGT TTA TTA TAT GAT
 CAA TGG GTA TTA TCA CCT CCT CAT AAA GAA AGA GTT AAT CAT TTA GGT AAT TTA GIA
 ATT ACC TGG GCA GCT CAA ACT ATT AAA CAT CAA GCT ATT AAA TTA GCT AAC TTA TTT
 ATT GTA ATT AAA AAA CTT ATT CCT ATT AAT TTA GTT GAA AAT TAT TTA ACA CCT ATG
 AGT CTG GCA TAT TGG ATT ATG GAT GGA GGT AAA TGG GAT TAT AAA AAT TCT CTT
 AAT AAA AGT ATT GTA TTA AAT ACA AGT ATT ACT TTT GAA GAA TAT TTA CTT
 AAA GGT TTA AGA AAT AAA TTT CAA TTA AAT TGT ATT ATT AAA AAT AAA CCA
 ATT ATT TAT ATT GAT TCT AGT AGT ATT TTT TAT TTA ATT AAA CCT ATT TTA
 ATT CCT CAA ATG ATT TAT AAA CTG CCT AAT ACT ATT TCA TCC GAA ACT ATT TTA AAA TAA

FIG. 1

Bam HI

CCGGATCCATG CAT ATG AAA AAC ATC AAA AAC CAG GAA ATG AAC CTG GGT CCG AAC TCT
 AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC ATC GAA CAG TTC GAA GCA
 K L L K E Y K S Q L I E L N I E Q F E A
 1. GGT ACT GGT CTG ATC CTG GGT GAT **GAT** GCT TAC ATC CGT TCT CGT GAT GAA GGT AAA ACC TAC
 G I G L I L G A Y R S R D E G K T Y
 TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC GTA TGT CTG CTG TAC GAT
 C M Q F E W K N K A Y M D H V C L L Y D
 CAG TGG GTA CTG TCC CCG CCG CAC AAA AAA GAA CGT GTT AAC CAC CTC GGT AAC CTG GTA
 Q W V L S P H K K E R V N H L G N L V
 ATC ACC TGG GGC GGC CAG ACT TTC AAA CAC CAA GCT TTC AAC AAA CTC GCT AAC CTG TTC
 I T W G A Q T F K H Q A F N K L A N L F
 ATC GTT AAC AAC AAA ACC ATC CCG AAC AAC CTG GTT GAA AAC TAC CTG ACC CCG ATG
 I V N K K T I P N N L V E N Y L T P M
 2. TCT CTG GCA TAC TGG TTC ATG **GAT** GAT GGT GGT AAA TGG GAT TAC AAC AAA AAC TCT ACC
 S L A Y W F M **D** G K W D Y N K N K T
 AAC AAA TCG ATC GTC CTG AAC ACC CAG TCT TIC ACT TTC GAA GAA TAC CTC GGT
 N K S I V L N T Q F S F E V E Y L V
 AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC CTA AAA ATC AAC AAA ACC CCG
 K G L R N K F Q L N C Y K I N K N K P
 ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC CTG ATC AAA CCG TAC CTG
 I Y D S M S Y L I F Y L K P Y L
 ATC CCG CAG ATG ATG TAC AAA CCT CGG AAC ACT ATC TCC TCC GAA ACT TTC CTG AAA TAA
 I P Q M M Y K L P N T I S S E T F L K
 TAAGTCGACTGCAGGATCCGGTAAGTAAGTAA
 Sall PstI BamHI

1 and 2: THESE AMINO ACIDS ARE ABSOLUTELY NECESSARY TO PRODUCE CATALYTIC ACTIVITY OTHER SUBSTITUTIONS ARE POSSIBLE, SUCH AS DELETIONS OF THE 10 FIRST AMINO ACIDS.

FIG. 2

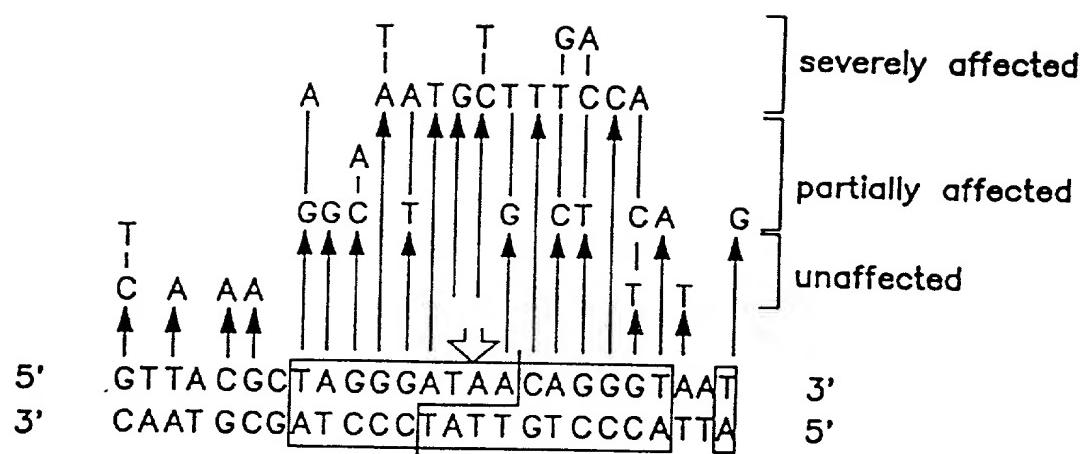


FIG. 3

1800
1800

FIG. 4A

2544	ACT CCC CAT CCC CCT GTT GAC AAT	P H	V	TCATGGCTCGTATA ATG TGT GGA ATT GTG AGC GGA	2606
2607	TAA CAATTACACAGAACAGGATCC	BamHI		ATG CAT ATG AAA AAC ATC AAA AAC CAG GTA ATG	2670
8				M H M K N I K N Q V M	12
2671	AAC CTG GGT CCG AAC TCT AAA CTT			CAG TCC AAA TAC GAA GAA CTG ATC GAA CTG AAC	2730
13	N L G P N S K			L K E Y K S I E L	32
2731	ATC GAA CAG TTC GCA GGA GGT ATC	F E A G		CTG GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT	2790
33	Q F Q E			I L G D A Y R S R	52
2791	GAT GAA GGT AAA ACC TAC TGT ATG	M C Y C		Q F E W K N K A Y M D H	2850
53	D E G K			M D H	72
2851	GTA TGT CTG CTG TAC GAT CAG TGG GTA	V D Q L Y		P H S P H K K A Y E R V	2910
73				P H K K A Y E R V	92
2911	CAC CTG GGT AAC CTG GTA ATC ACC	H N G		Q G A C T T C A A C C A C G A C T T C A A C	2970
93				T W G A Q T F K H Q A F	112
2971	AAA CTG GCT AAC CTG TTC ATC GTT	A F N		N K K V W F M N N L V	3030
113				A C C T C G A A C A A A A A C T C C G A A C A A C	132
3031	AAC TAC CTG ACC CCG ATG TCT	Y T P M		Y W F M D D G G T G A T G A T G G T A A A T G G	3090
133				A A C T C G A C C A A C A A C T C G T C T T C A C T	152
3091	TAC AAC AAA AAC TCT ACC AAC AAA TCG ATC	Y N S T		Y L N T Q S F T F G A A	3150
153				T N K S I V	172
3151	GAA GTA GAA TAC CTG GTT AAG GGT	E Y K V		R N K F Q L C T G A A C T T C C A A C T G T A C	3210
173				Y D S M Y C Y V K	192
3211	ATC AAC AAA AAC AAA CCG ATC ATC TAC	N K P I		T G T C T A C T G T C T A C T G T C T A C	3270
193				Y I D S M Y C Y V K	212
3271	CTG ATC AAA CCG TAC CTG ATC CCG CAG ATG TAC	I K P Y L		Q M M Y K L P N T S T C C T C C	3330
213				Y K L P N T S T C C T C C	232
3331	GAA ACT TTC CTG AAA TAA	E F L K		Satt PstI TAAGTCGACCTGGAGCCCCAAGCTGGACTGGCCGTTACAACGTCGTGACT	3404
233					238

FIG. 4B

-2	M	H	M	K	N	I	5	K	K	N	Q	V	10	M	N	L	G	P	N	S	
			20											30							
K	L	L	K	E	Y	K	S	Q	L	I	E	L		N	I	E	Q	F	E	A	
G	I	G	L	I	L	G	D	A	Y	I	R	S	50	R	D	E	G	K	T	Y	
C	M	Q	F	E	W	K	N	K	A	Y	M	D	70	H	V	C	L	L	Y	C	
Q	W	Y	80	L	S	P	P	H	K	K	E	R	Y	90	N	H	L	G	N	L	Y
I	T	W	100	G	A	Q	T	F	K	H	Q	A	F	110	N	K	L	A	N	L	F
I	V	N	120	N	K	K	I	I	P	N	N	L	V	130	E	N	Y	L	T	P	M
G	L	A	140	Y	W	P	M	D	D	G	G	K	W	150	D	Y	N	K	N	S	I
N	K	S	160	I	V	L	N	T	Q	S	F	T	F	170	E	E	V	E	Y	L	V
K	G	L	180	R	N	K	F	Q	L	N	C	Y	V	190	K	I	N	K	N	K	P
I	I	Y	200	I	D	S	M	S	Y	L	I	F	Y	210	N	L	I	K	P	Y	L
I	P	Q	220	M	M	Y	K	L	P	N	T	I	S	230	S	E	T	F	L	K	*

Positions that can be changed without affecting enzyme activity (demonstrated)
 positions -1 and -2 are not natural. The two amino acids are added due to cloning strategies

positions 1 to 10: can be deleted
 position 36: G is tolerated
 position 40: M or V are tolerated
 position 41: S or N are tolerated
 position 43: A is tolerated
 position 46: V or N are tolerated
 position 91: A is tolerated
 positions 123 and 156: L are tolerated
 position 223: A and S are tolerated

Changes that affect enzyme activity (demonstrated)

position 19: I to S
 position 38: I to S or N
 position 39: G to D or R
 position 40: L to Q
 position 42: L to R
 position 44: D to E G or H
 position 45: A to E or D
 position 46: Y to D
 position 47: I to R or N
 position 80: L to S
 position 144: D to E
 position 145: D to E
 position 146: G to S
 position 147: G to S

FIG. 5

Group I Intron Encoded Endonucleases and Related Endonucleases

	ENDONUCLEASE	RECOGNITION SEQUENCE	CLEAVAGE SITE	▽ INTRON SITE
TWO DODECAPEPTIDE FAMILY (OR 4 BP CUTTERS)	I-Sce I (<i>Saccharomyces</i> mitochondria)	CGC TAGGGATAACAGGGTAATATAGC GCG ATCCCTATTGTCCCATTATATCG	▽	
	I-Sce IV (<i>Saccharomyces</i> mitochondria)	TTCTCATGATTAGCTCTAATCCATGG AAGAGTACTAATCGAGATTAGGTACC	▽	
	I-Sce II (<i>Saccharomyces</i> mitochondria)	C TTGGTCATCCAGAAGTATATATT GAAACCAGTAGGTCTTCATATAAAA	▽	
	I-Ceu I (<i>Chlamydomonas</i> chloroplast)	TAACGGTCCTAACGGTAGCGAAATTCA ATTGCCAGGATTCCATCGCTTAAGT	▽	
	I-Ppo I (<i>Physarum</i> nucleus)	TGACTCTCTAACGGTAGGCCAAATGCC ACTGAGAGAATTCCATCGGTTACGG	▽	
	I-Sce III (<i>Saccharomyces</i> mitochondria)	GGAGGTTTTGGTAAC TATTATTATTAC CCTCCAAAAACCATGATAAATAATGG	▽	
	I-Cre I (<i>Chlamydomonas</i> chloroplast)	GGGTTCAAAACGTCTGAGACAGTT CCCAAGTTTGCA GCACTCTGTCAA	▽	
	Endo. Sce I(RF3) (<i>Saccharomyces</i> mitochondria) (Non intronic)	GATGCTGTAGGCATAGGCTTGGTTAT CTACGACATCCGTATCCGAACCAATA		
	HO (<i>Saccharomyces</i> nucleus) (Non intronic)	CTT TCCGCAACAGTATAATT TATAAA GAAAGGCGTTGTCA TATTAAAATATT		
	I-Csm I (<i>Chlamydomonas</i> mitochondria) (Putative endonuclease)	ACCATGGGGTCAAATGTCTTCTGGG TGGTACCCAGTTACAGAAAGACCC	▽	
OTHER STRUCTURAL FAMILIES	I-Pan I (<i>Podospora</i> mitochondria) (Putative endonuclease)	GTGCCTGAATGATATT TATTACCTT CACGGACTTACTATAAATAATGGAAA	▽	
	(Bacteriophage T4)			
	I-Tev I	CAACGCTCAGTAGATGTTTCTTGGGTCTACCGTTAA GTTGCGAGTCATCTAACAAAGAACCCAGATGGCAAATT	▽	
	I-Tev II	CAAGCTTATGAGTATGAAGTGAACACGTATT GTTCGAATACTCATACTTCACCTTGTCATAA	▽	
	I-Tev III	GCTATT CGTTT TATGTATCTTTGC GTGTAGCTTAA CGATAAGCAAAAATACATAGAAAACGCCACATCGAAATT	▽	

FIG. 6

EXPRESSION VECTORS

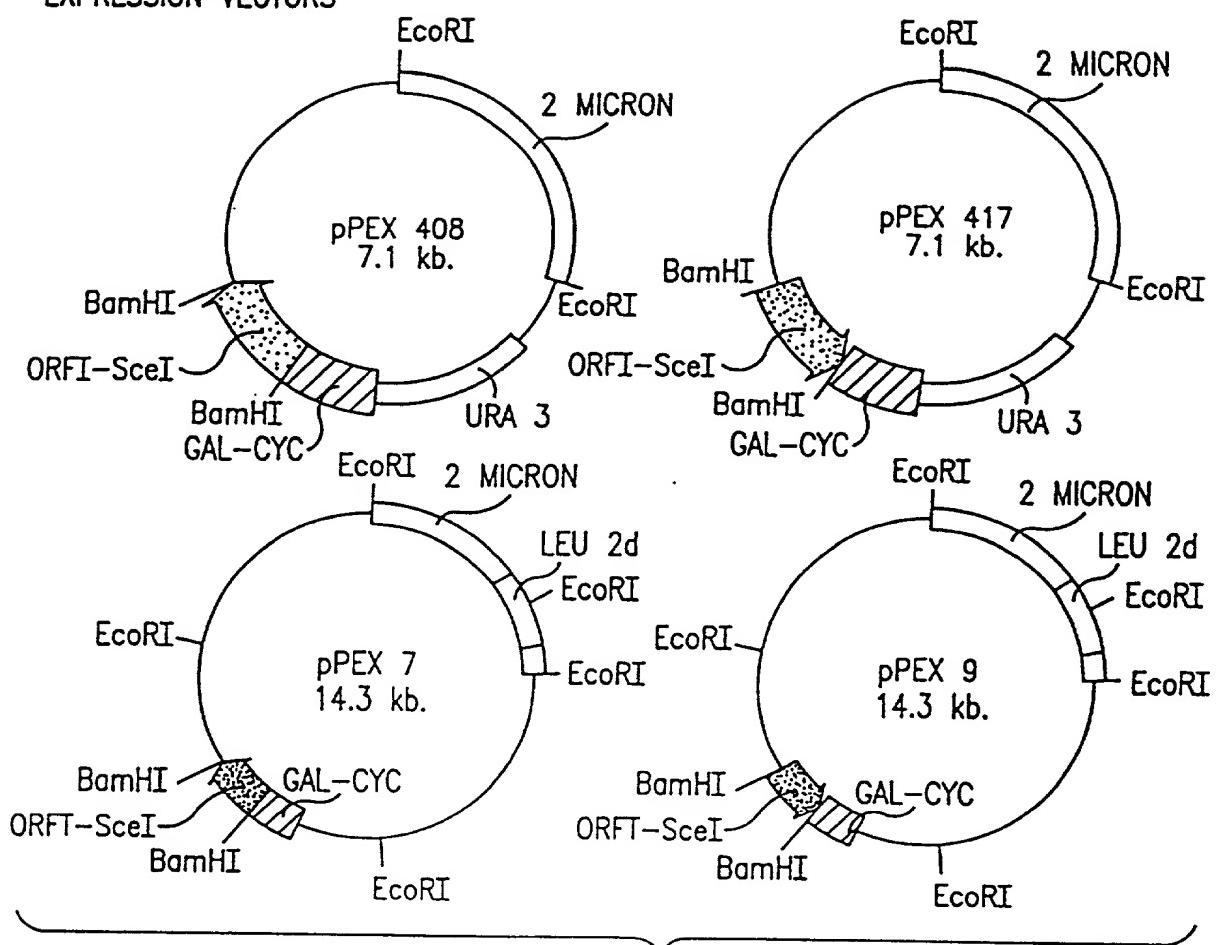


FIG. 7

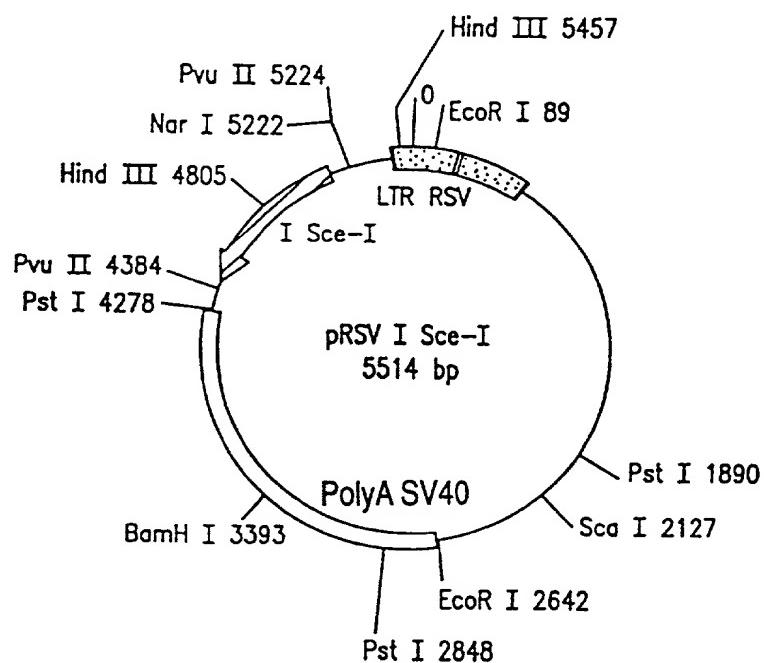


FIG. 8

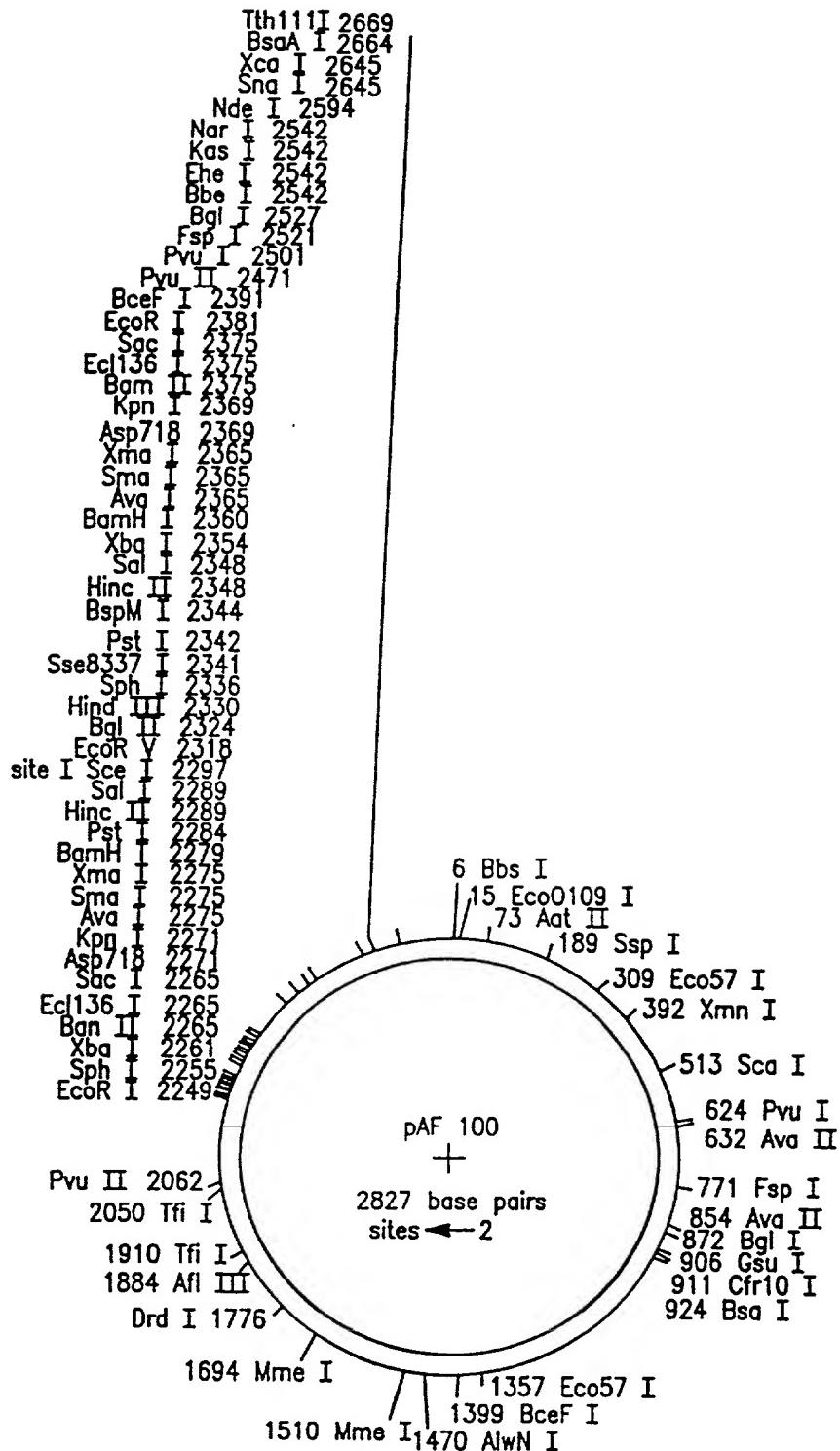


FIG. 9

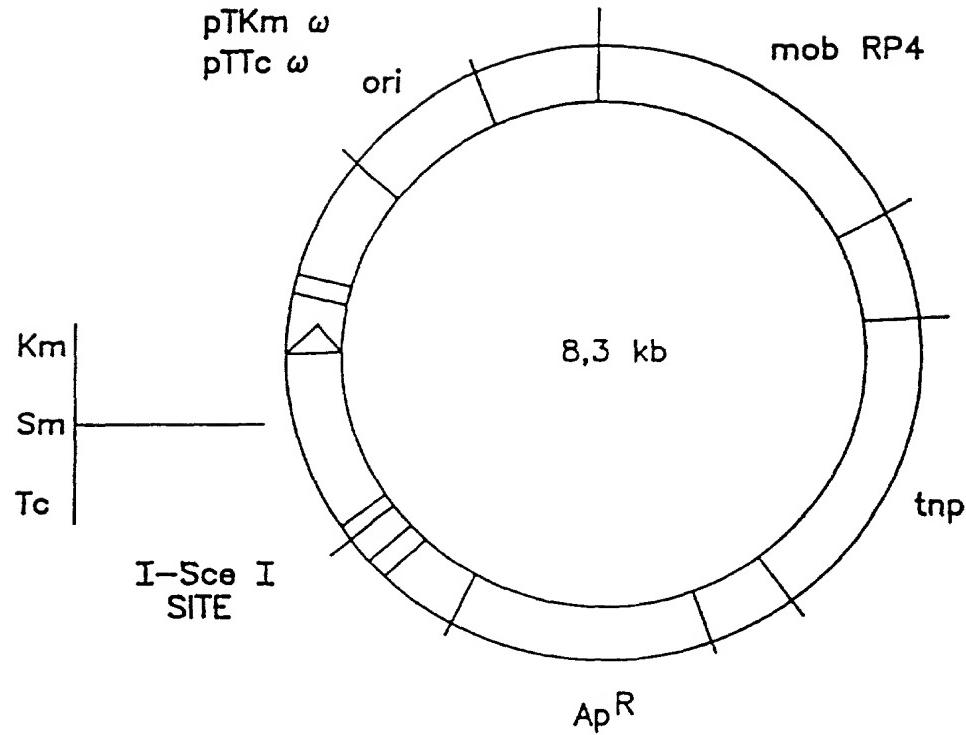
Sau3A I
 Mbo I
 Dpn II
 ScrF I
 Nci I
 Msp I
 Hpa II
 Dsa V
 BstK I
 Xma I
 Sma I
 ScrF I
 Nci I
 Dsa V
 Rsa I Dpn I
 Nla IV Nla IV
 Sac I Csp6 I Alw I
 HgiA I BstK I Taq I
 Ec1136 I BsaJ I Sal I
 Sph I Rma I Kpn I BstY I Hin c II
 NspC I Bsp1286 I Bcn I Sfe I
 EcoR I Xba I Ban I BamH I Acc I I-Sce I
 Apo I Nla III Ban II Bcn I Sfc I Rma I
 Taq I Nsp7524 I Alu I Ava I Pst I Hga I
 Alu I Nsp I Bfa I Asp718 Alw I BspW I Bfa I EcoR V
 CCAAGCTCGAATTCGATGCTCTAGAGCTCGGTACCCGGGATCCTGCAGTCGACGCTAGGGATAACAGGGTAATACAGAT 2320
 GGTCGAGCTTAAGCGTACGAGATCTCGAGCCATGGGCCCTAGGACGTCAGCTGCCATCCCTATTGTCCCATTATGTCTA
 2244 2255 2262 2271 2279 2286 2296 2318
 2247 2255 2266 2275 2284 2292
 2249 2256 2265 2275 2284 2296
 2249 2261 2271 2279 2289 2297
 2255 2265 2276 2284
 2255 2262 2271 2279 2289
 2265 2275 2289
 2265 2275 2290
 2265 2272 2280
 2271 2279
 2272 2280
 2275
 2275
 2275
 2275
 2275
 2275
 2276
 2276
 2276
 2276
 2276
 2276
 2280
 2280
 2280

FIG. IOA

ScrF I
 Nci I
 Msp I
 Hpa II
 Dsa V
 BstK I
 Bcn I
 Xma I
 Sma I
 ScrF I
 Nci I
 Dsa V
 BstK I
 BsaJ I
 Sau3A I
 Mbo I Sac I
 Dpn II HgiA I
 Alu I Rsa I
 Hind III Sfc I Ple I Dpn I
 Dde I Pst I Hinf I BstY I Nla IV Csp6 I
 Sau3a I Sse8337 I BarnH I Kpn I EcoR I
 Mbo I Nla III Taq I Alw I Ban I Alu I BceF I
 Dpn II Sph I Sal I Rma I Bcn I Ec1136 I Hae III
 Dpn I NspC I Hinc II Mnl I Ava I Bsp1286 I Gdi II
 BstY I Nsp7524 I Acc I Bfa I BsaJ I Ban II Eae I
 Bgl II Nsp I BspM I Xba I Alw I Asp718 Taq I Bsr I
 ATCAGATCTAACGTTGCATGCCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCACTGGCCGTCGTTT 2400
 TAGCTAGATTGAAACGTACGGACGTCCAGCTGAGATCTCCTAGGGGCCATGGCTCGAGCTTAAGTGACCGGGCAGCAA
 2324 2336 2344 2354 2361 2369 2379 2387
 2324 2336 2348 2355 2364 2375 2389
 2325 2336 2348 2358 2365 2375 2389
 2325 2336 2348 2355 2365 2375 2390
 2325 2337 2349 2360 2369 2376 2391
 2325 2341 2360 2369 2381
 2328 2342 2351 2360 2369 2381
 2330 2342 2351 2360 2370
 2331 2342 2361 2370
 2361 2375
 2361 2375
 2361
 2365
 2365
 2365
 2365
 2365
 2365
 2365
 2366
 2366
 2366
 2366
 2366
 2366
 2366
 2366
 Mae II
 ScrF II
 EcoR II
 Dsa V
 BstN I
 BstK I
 Mae II Bsr I
 BsaJ I
 Mse I Fnu4H I Fok I Alu I
 Bbv I Pvu II
 NspB II
 TACAACGTCGTGACTGGGAAAACCCCTGGCTTACCCAACCTTAATGCCCTTGAGCACATCCCCCTTCGCCAGCTGGCGT 2480
 ATGTTGCAGCACTGACCCTTTGGGACCGCAATGGGTGAATTAGCGGAACGTCGTAGGGGGAAAGCGGTCGACCGCA
 2405 2413 2423 2440 2451 2471
 2410 2424 2451 2471
 2424 2457 2472
 2424
 2424

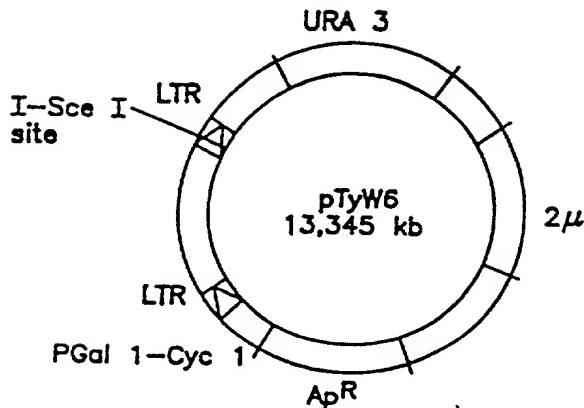
FIG. IOB

NAMES: pTSm ω
pTKm ω
pTTc ω



Construction: pGP 704 from De Lorenzo, with transposase gene and insertion of the linker [I-SceI] in NotI unique site

FIG. I I



Construction: pD 123, from J.D. Boeke
with insertion of a linker[I-SceI-NotI] in BamHI

FIG. 12

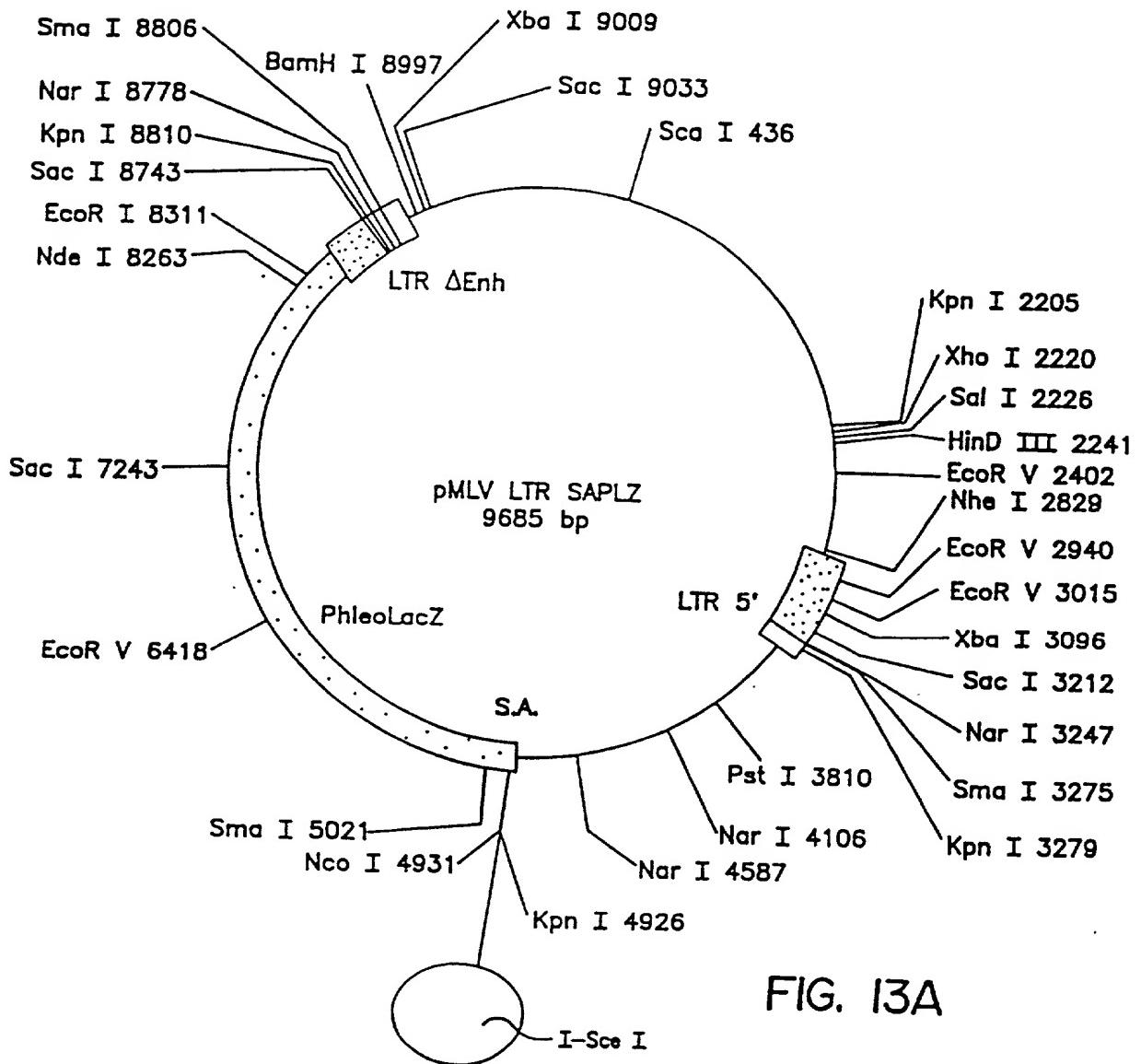


FIG. 13A

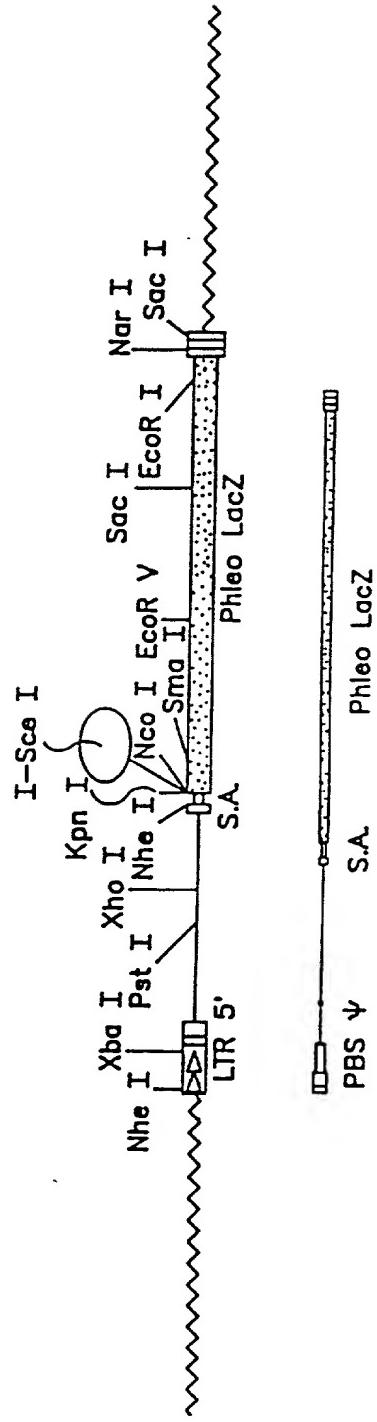


FIG. 13B

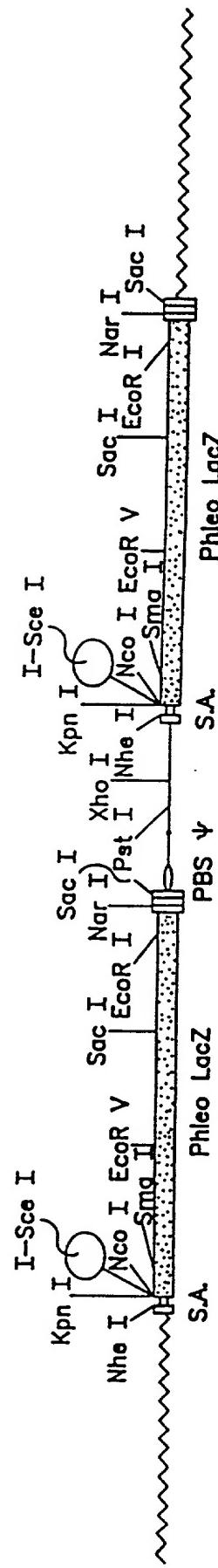
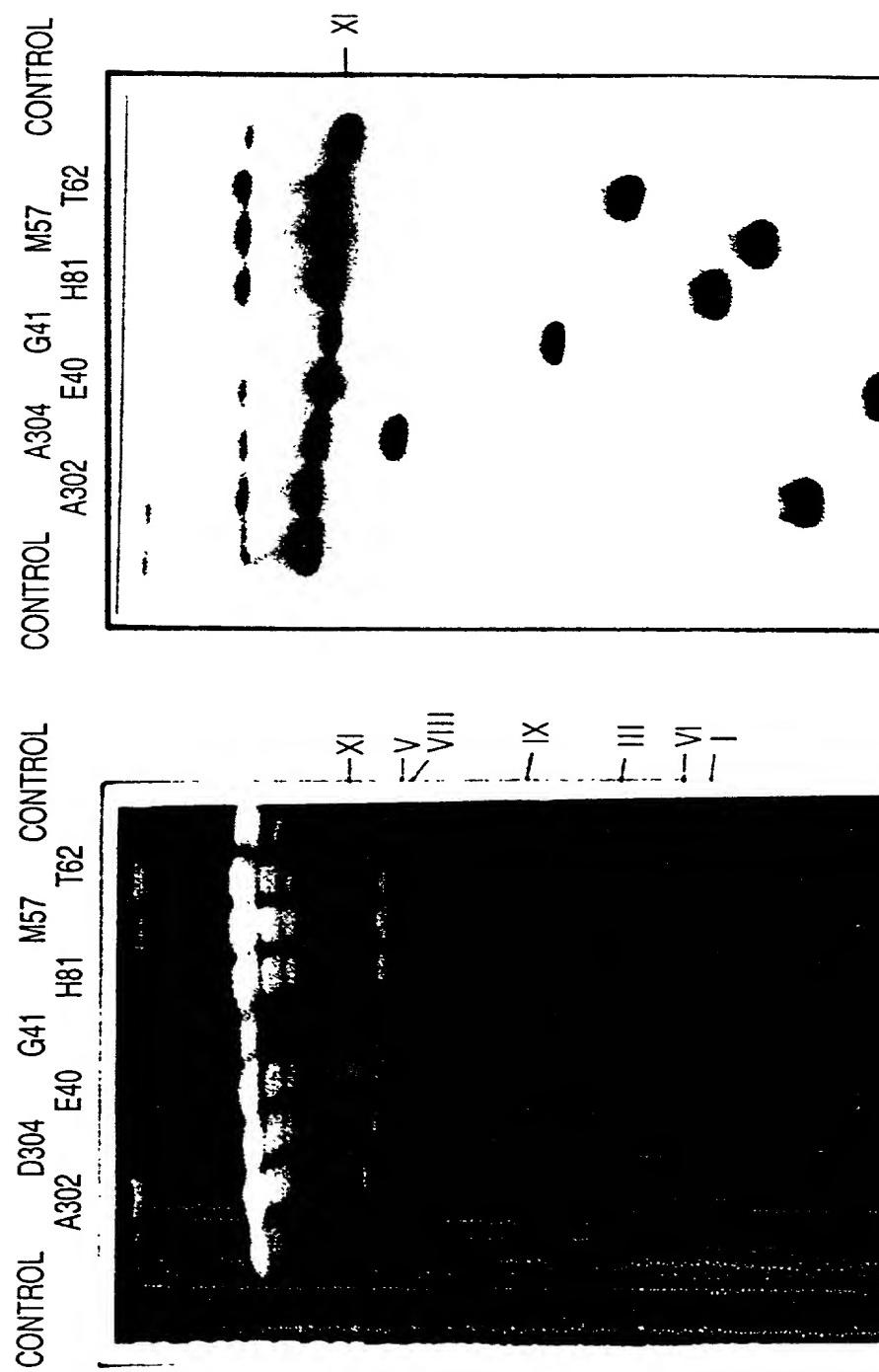


FIG. 13C



LEFT END PROBE
COSMID pUKG 040

FIG. 14B

FIG. 14A

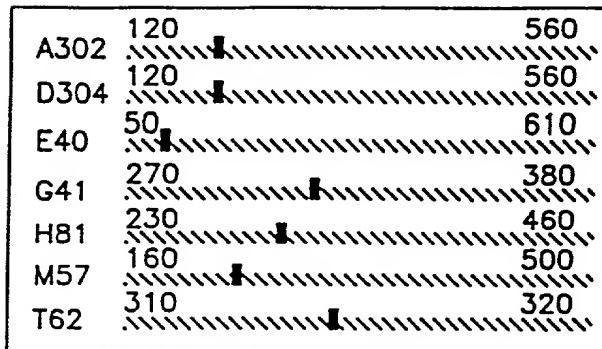


FIG. 15A

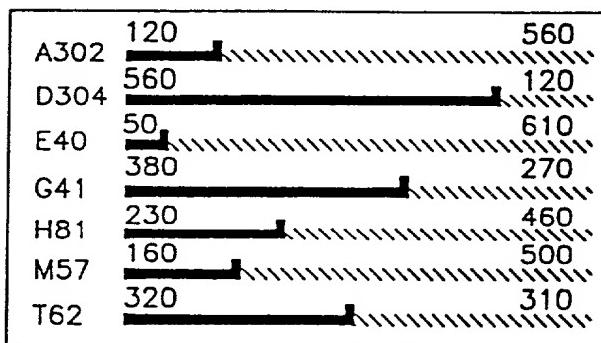


FIG. 15B

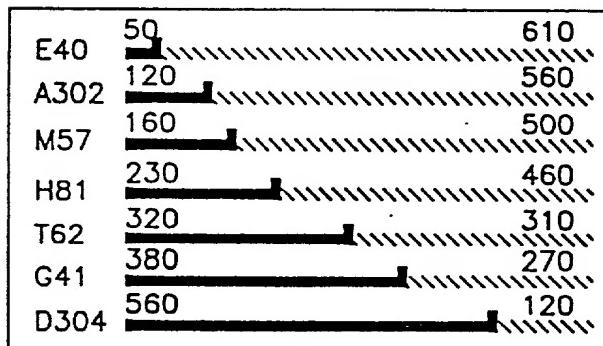


FIG. 15C

FIG. 15D

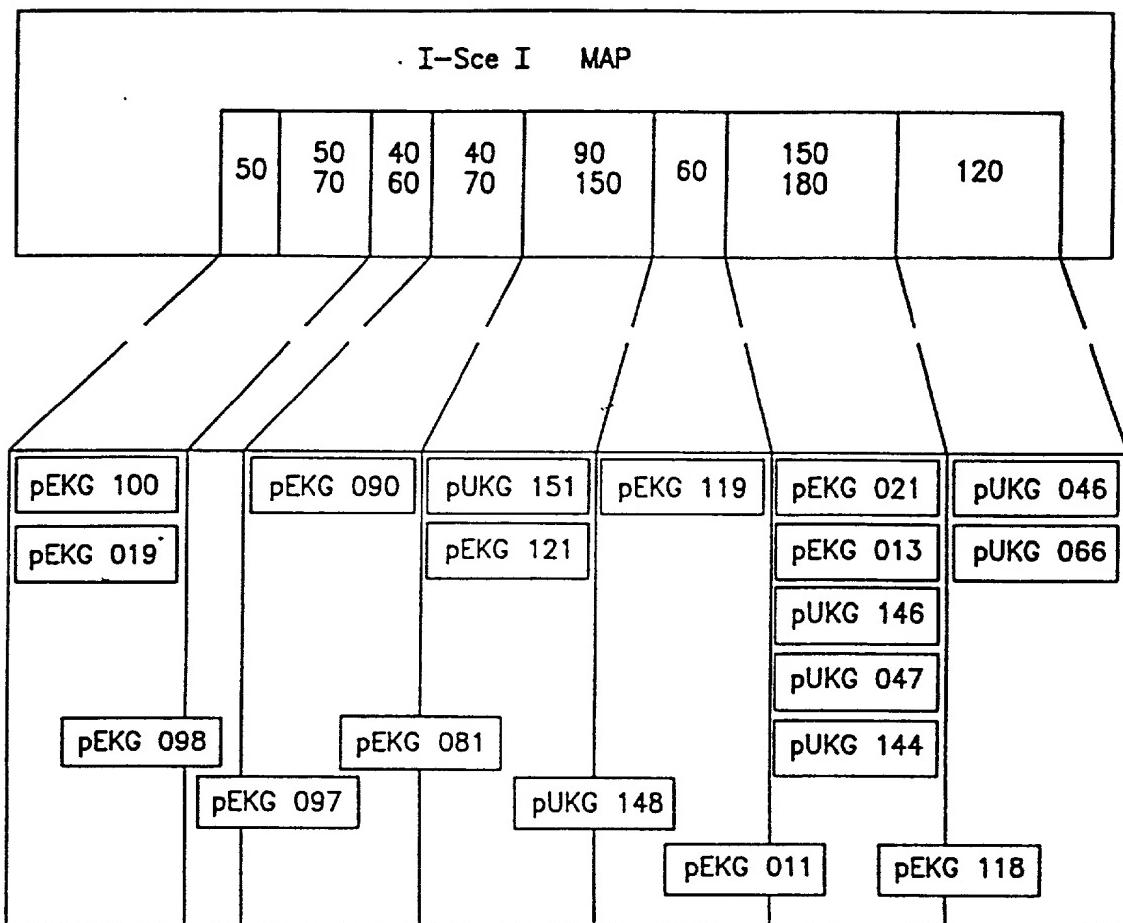


FIG. 15E

ESTIMATED SIZE 850

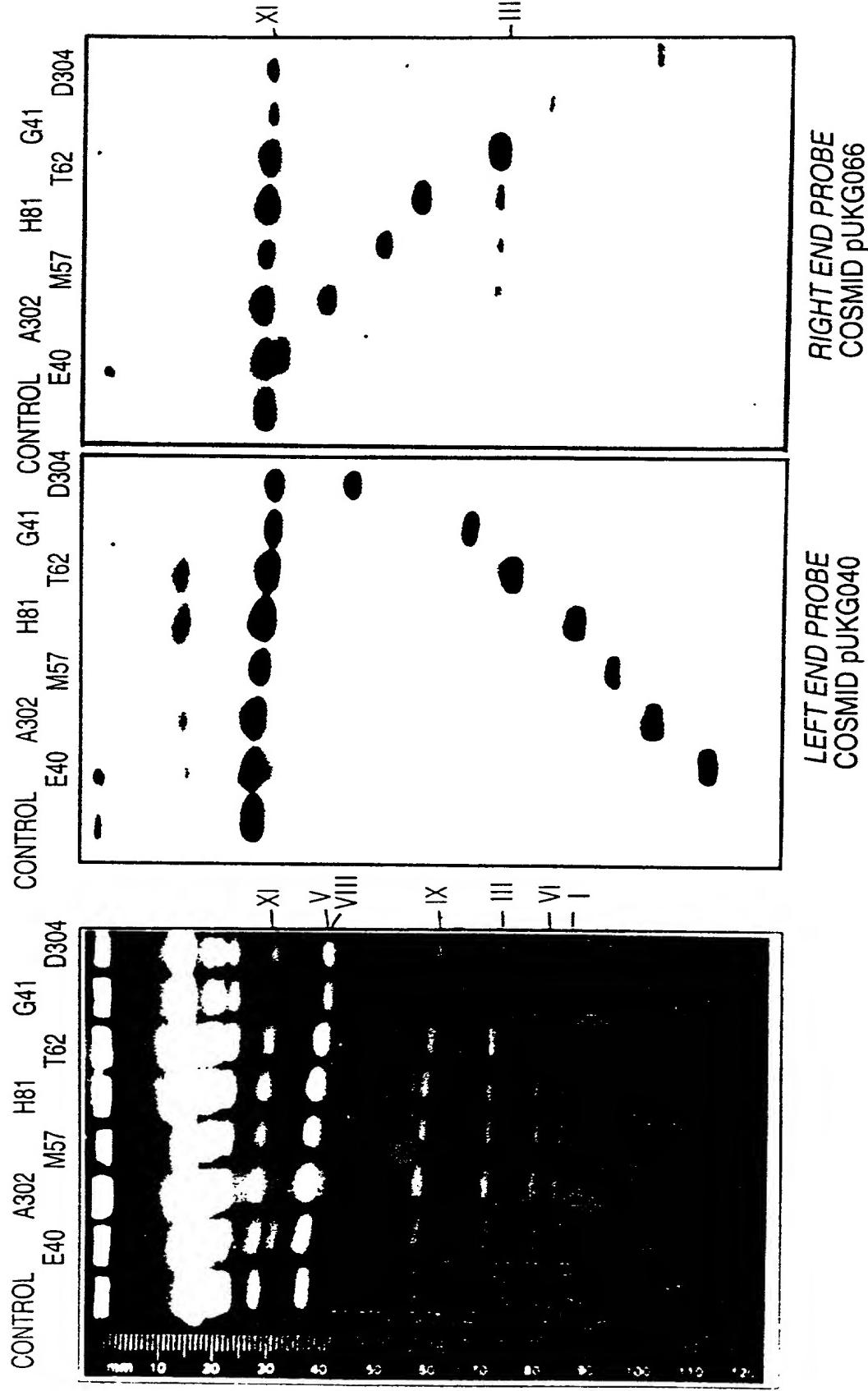


FIG. 16A

FIG. 16B

FIG. 16C

LEFT END PROBE
COSMID pUKG040

RIGHT END PROBE
COSMID pUKG066

卷之三

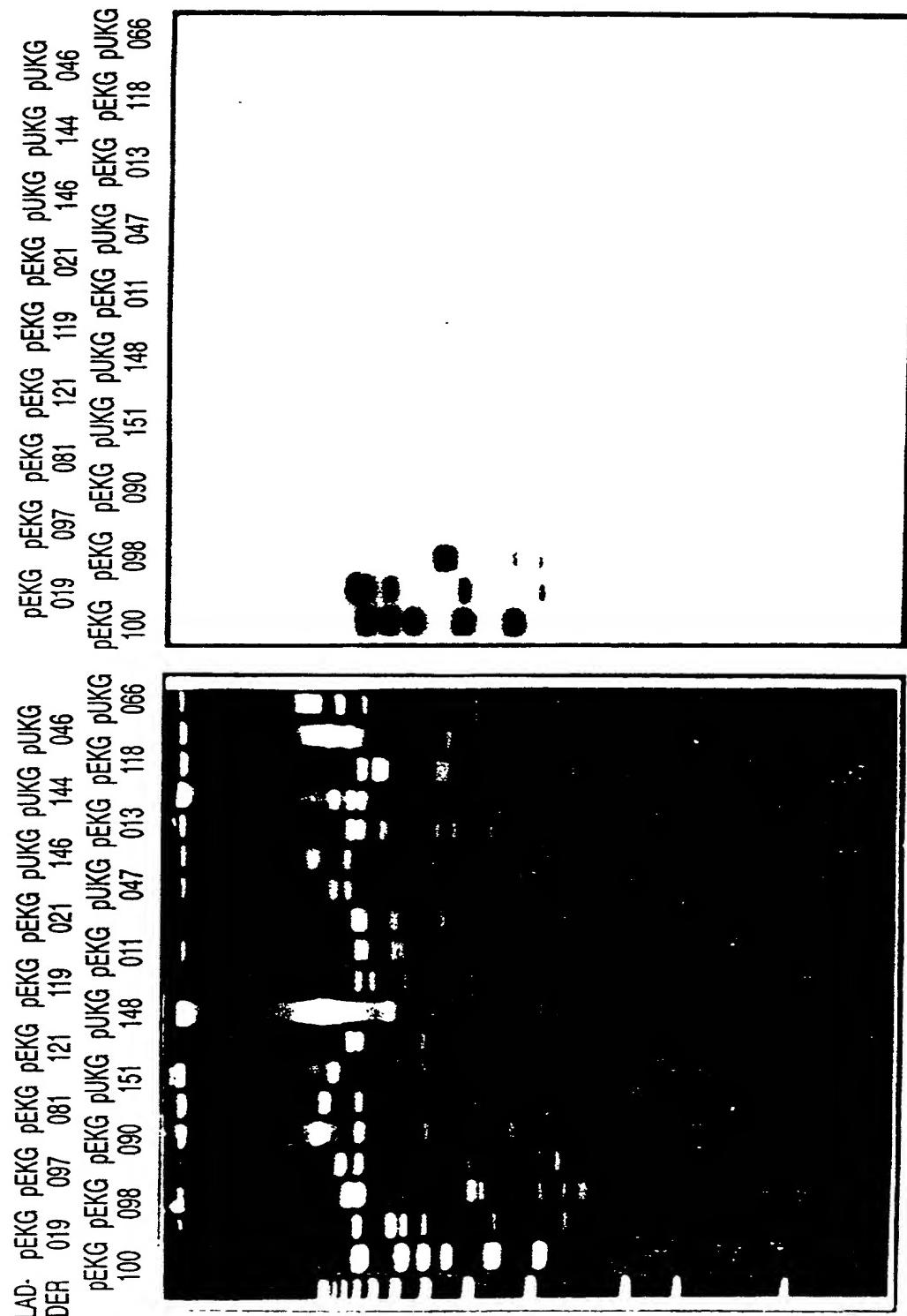


FIG. 17A

FIG. 17B

As a result, the number of species per genus was significantly higher in the *Leptospiraceae* than in the *Neurosporaceae* (Table 1).

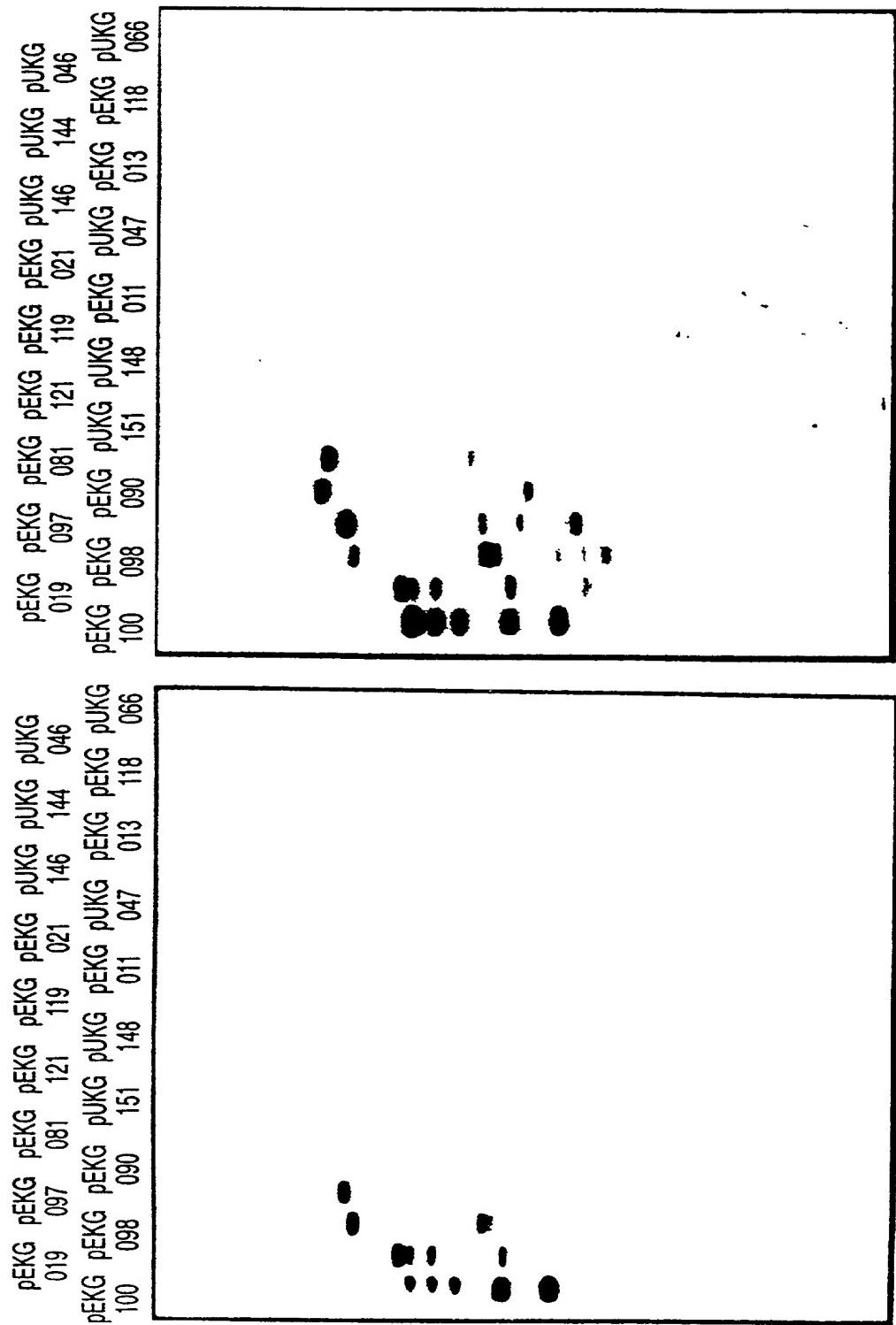


FIG. 17C

FIG. 17D

pEKG
019 097 081 121 119 021 146 046 019 097 081 121 119 021 146 144 046
pEKG
100 098 090 151 148 011 047 013 118 066 100 098 090 151 148 011 047 013 118 066

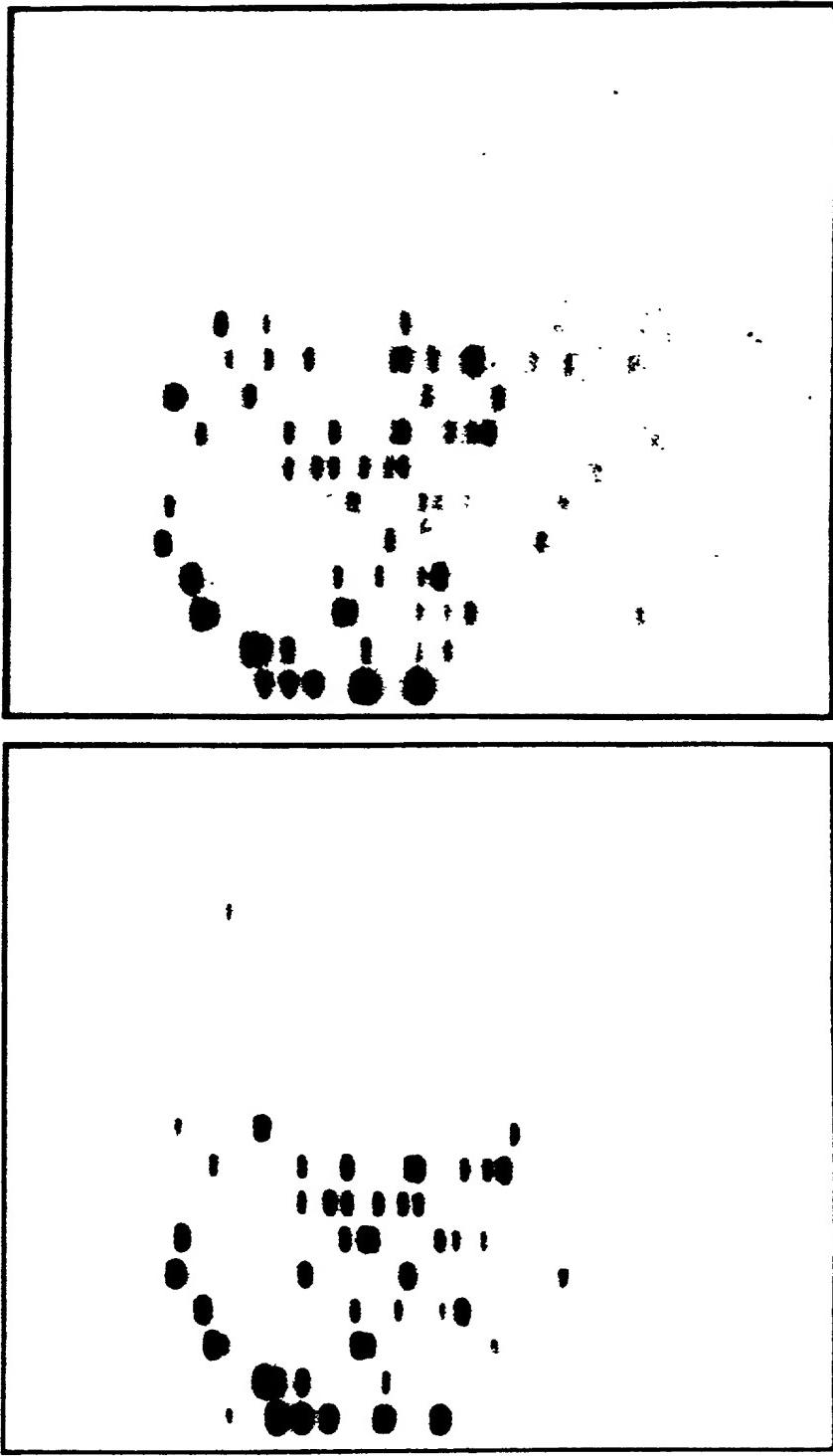


FIG. 17E

FIG. 17F

pEKG pEKG pEKG pEKG pEKG pEKG pEKG
019 097 081 121 119 021 146 046
pEKG pEKG pEKG pEKG pEKG pEKG pEKG
100 098 090 151 148 01 047 013 118 066

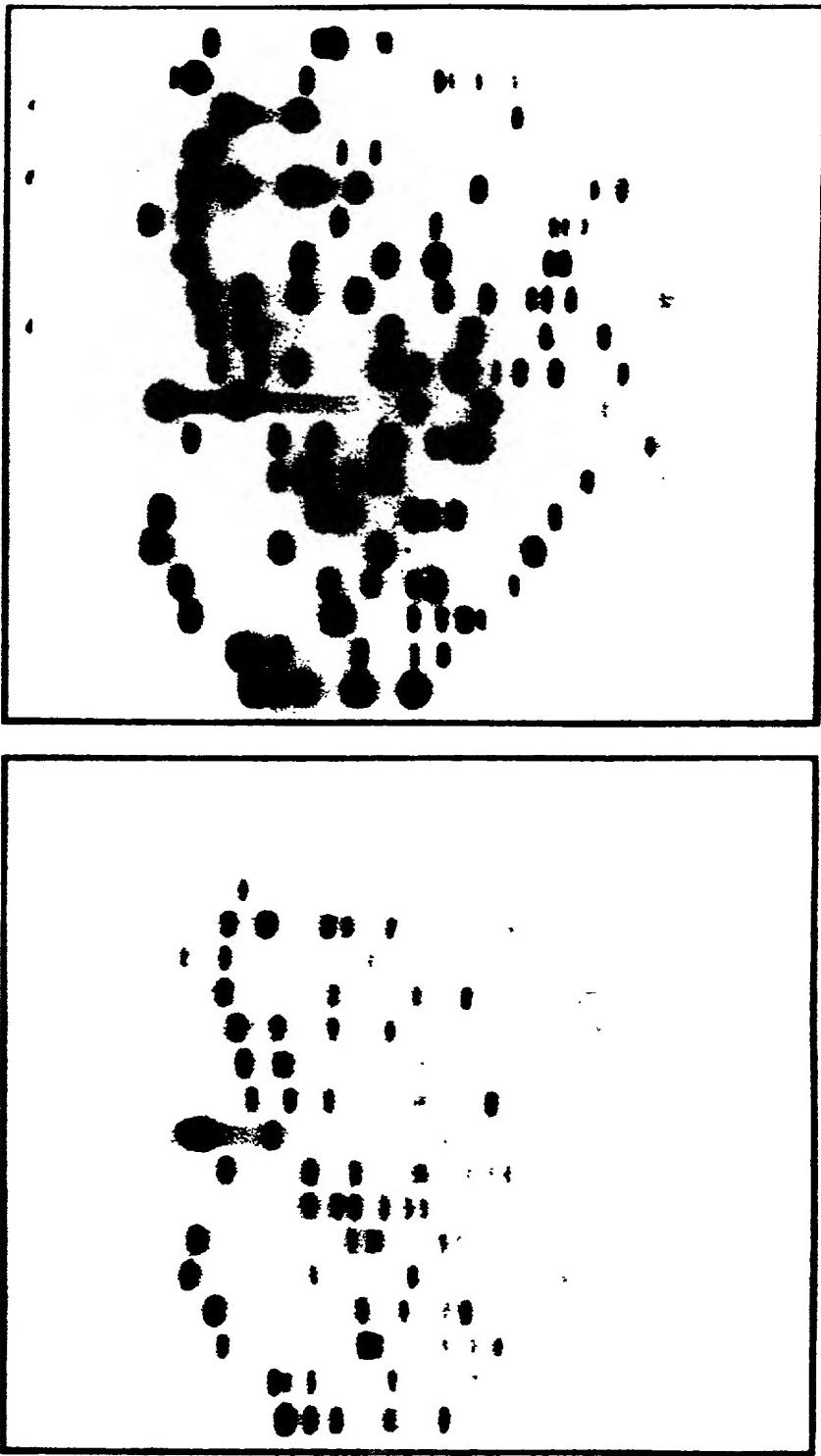


FIG. 17G

FIG. 17H

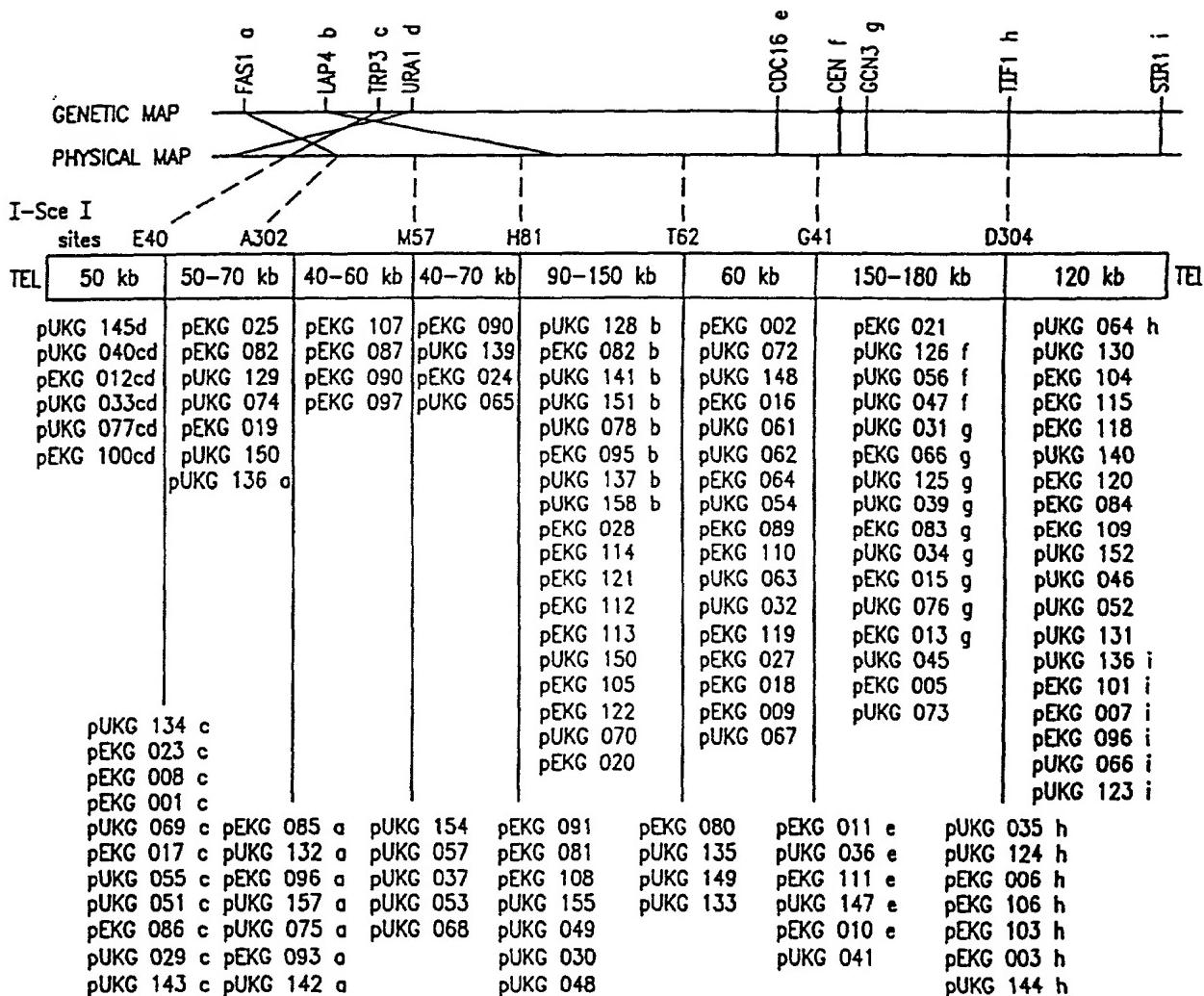


FIG. 18

103740 - 0375260

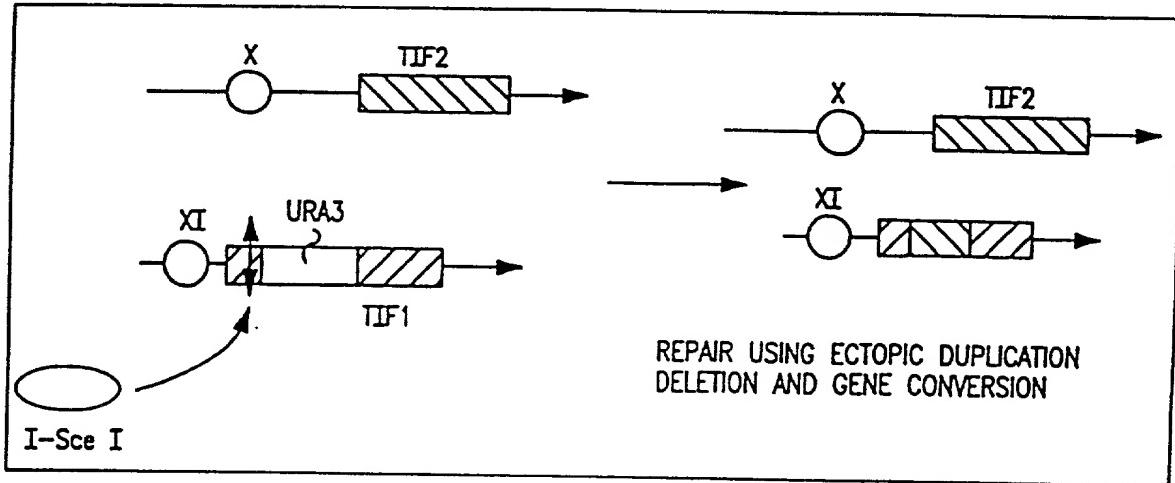


FIG. 19A

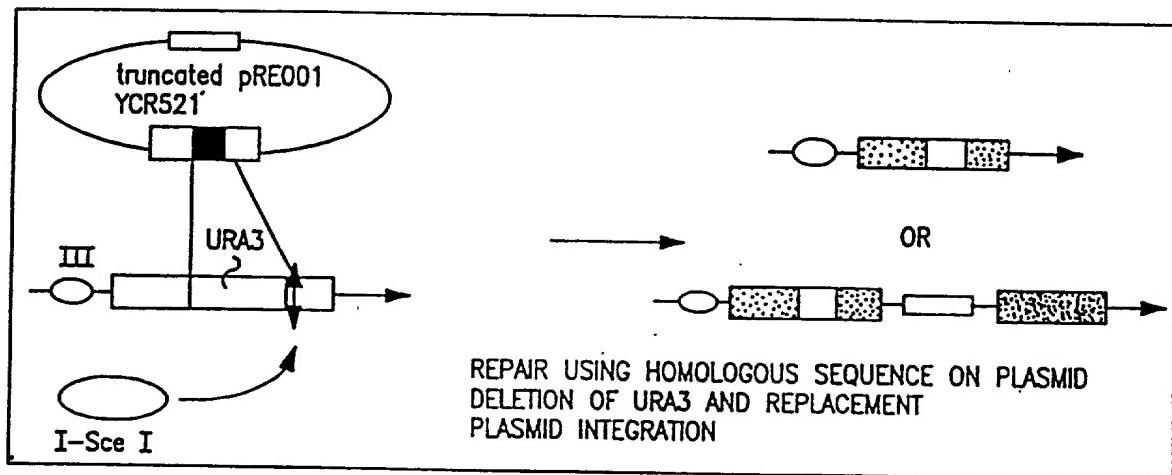


FIG. 19B